

EGU2020-20132, updated on 19 Jan 2021

<https://doi.org/10.5194/egusphere-egu2020-20132>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



## **Towards synthetic data generation for machine learning models in weather and climate**

**David Meyer**

Department of Meteorology, University of Reading, Reading, United Kingdom of Great Britain and Northern Ireland

The use of real data for training machine learning (ML) models are often a cause of major limitations. For example, real data may be (a) representative of a subset of situations and domains, (b) expensive to produce, (c) limited to specific individuals due to licensing restrictions. Although the use of synthetic data are becoming increasingly popular in computer vision, ML models used in weather and climate models still rely on the use of large real data datasets. Here we present some recent work towards the generation of synthetic data for weather and climate applications and outline some of the major challenges and limitations encountered.